LEVEL II

TR-81-39-5
November 30, 1981

AD A109326

TRAINING REQUIREMENTS ANALYSIS FOR TACTICAL DOMAINS

FINAL REPORT: 1 June 1979 to 30 September 1981

Prepared for
The Air Force Office of Scientific Research (AFSC).
Contract Number AFOSR-F49620-79-C-0179

KLEIN ASSOCIATES
740 Wright Street • Yellow Springs, Ohio 45387 • Phone (513) 767-2691

82 01 04 057

TR-81-39-5

TRAINING REQUIREMENTS ANALYSIS FOR TACTICAL DOMAINS

FINAL REPORT: 1 June 1979 - 30 September 1981

Prepared by

Gary A. Klein, Ph.D.

November, 1981

KLEIN ASSOCIATES
740 Wright St.
Yellow Springs, Ohio 45387
(513)767-2691

Prepared for the Air Force Office of Scientific Research (AFSC), under Contract Number AFOSR-F49620-79-C-0179.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR-81-39-5	2. GOVT ACCESSION NO. AD-A109 326	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Training Requirements Analysis for Tactical Domains		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report 6/1/79 - 9/30/81
		6. PERFORMING ORG. REPORT NUMBER see 1.
7. AUTHOR(s) Gary A. Klein		8. CONTRACT OR GRANT NUMBER(s) AFOSR-F49620-79-C-0179
9. PERFORMING ORGANIZATION NAME AND ADDRESS Klein Associates 740 Wright Street Yellow Springs, Ohio 45387		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2313/A2
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling AFB, DC 20332		12. REPORT DATE November 30, 1981
		13. NUMBER OF PAGES 40
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

analogical reasoning	proficiency	decision aids
analogical inference	complex learning	
reasoning by analogy	prediction	
reasoning	hypothesis generation	
expertise	decision making	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Abstract

DD FORM 1 JAN 73 1473

Unclassified 394586

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

1. Abstract

The goal of this research program was to develop and test a model of proficient performance. Three aspects of proficient performance were studied: recognitional capacities, perceptual learning, and the use of analogies. → Pay 2

Weitzenfeld (1977, 1981a) demonstrated on theoretical grounds that knowing how to perform a task depends on recognitional capacity and cannot be reduced to knowledge of rules or procedures. This theoretical analysis complements the empirical demonstration of the existence of special recognitional capacities in experts, notably by Simon and colleagues (1973, 1980). We examined the hypothesis that the possession of these recognitional capacities is the driving force in the superiority of chess masters over chess experts (a lower category of ability). We did this by comparing the quality of moves in regulation time games and 5-minute games. Increased time should allow more detailed analyses of moves, but should have little effect on the move alternatives first recognized. Both masters and experts showed better performance with more time, indicating that they are both using calculational processes. However, the superiority of the masters did not increase with additional time; the masters were better for the 5-minute games, and maintained the same level of superiority for the regulation games. This suggests that both masters and experts show the same level of calculational skills, and supports the hypothesis that the strength of the chess masters is primarily due to recognitional capacity --the ability to recognize the strongest option.

The previous work established the relevance of recognitional capacities. The growth of such discriminative abilities has been the focus of work by E. Gibson (1969). We hypothesized that more proficient

subjects have learned to use a greater number of dimensions for perceiving a task. This was studied for the Cardio-Pulmonary Resuscitation (CPR) skills of beginners, instructors, and paramedics (ten subjects in each group). We were able to pinpoint perceptual dimensions that could be used by paramedics, but not by lesser-skilled personnel. These differences could be used as the basis for defining perceptual training requirements for complex tasks. A second study applied the same paradigm to computer programming, comparing novices (1-3 years experience) with programmers (more than seven years of experience). Again, clear differences emerged in the use of perceptual dimensions.

It has been noted that chess masters can recall positions from a very large number of games and in chess analysis a position is characteristically compared to the same or similar positions reached in previous games through recent chess history. What role do such comparisons play in the proficiency of chess masters? How are analogical comparisons made and used?

We studied the use of analogical reasoning for generating predictions in technological environments. Three models of analogical reasoning were considered and rejected. The first is the standard a:b::c:d model employed by test-makers. The second stems from the use of analogical reasoning to generate new scientific hypotheses. Seven Air Force engineers were interviewed; all had used comparison cases as analogues for the task of predicting reliability of subsystems for the B-1. Neither model was able to account for the performance of the engineers. A third model claims that analogical reasoning is based on similarity matches, and is probabilistic. This model was rejected on conceptual grounds -- the processes it relies on are inadequate for the task. To replace these models, we developed a new theory of analogical reasoning, showing its basis in standard forms of

deductive logic. We also were able to define the conditions under which analogical reasoning will generate formally valid conclusions. This work is relevant for any area in which comparisons play an important role, such as the domain of technological improvement involving design changes in automobiles, aircraft, etc.

The research performed has implications for a number of applied areas, such as the development of methods for generating predictions under conditions of uncertainty, the design of programs for training personnel to reach high levels of proficiency, and the development of automated decision aids to support experienced C^2 personnel.

2. Research Objectives

The research obejctives of this program were to review the literature, perform theoretical analyses, develop research paradigms, and perform empirical research, in the area of highly proficient performance.

The nature of highly proficient performance has recently become the subject of a great deal of theoretical and empirical research. There seems to be general agreement that novices are learning individual steps, along with rules for when to perform these steps. However, highly proficient performance does not readily display characteristics of following steps, or rules. It has not been demonstrated that the behavior of experts can be defined in terms of computational operations on formally defined elements. This creates a challenge for the information processing approach to model highly proficient performance. It also creates an opportunity to examine some of the assumptions underlying the information processing approach in psychology, and to attempt to formulate alternative accounts of expertise that do not rely on a framework of computational operations. For applied purposes, the development of improved procedures for describing highly proficient performance could allow more effective methods for selecting and training highly competent personnel. Decision making is one example of a skill where our understanding of highly proficient performance can have important implications. The type of automated decision aids we can develop should be a function of the needs of proficient decision makers, rather than of the state-of-the-art in the relevant microprocessor technologies.

3. Problem Statement

Currently, there is no adequate theory of proficient performance. The thrust of psychological research on learning and competence until recently has been directed at novices acquiring an unfamiliar skill. However, it seems highly unlikely that the same processes will account for the difference between experts and novices. The assumption that experts are simply following the same procedures as novices (except that the experts are faster and more accurate) has not received empirical support.

Specifically, it may not be reasonable to assume, as information processing accounts do, that proficient performance of a skill depends on the ability to break tasks down into basic elements, to apply rules and procedures to these elements, and to use higher-level rules for the accomplishment of simpler rules. As we have discussed elsewhere (Klein, 1978), the postulation of basic elements of a task runs into the difficulties which led to the abandonment of logical atomism, and there are a variety of reasons to doubt that experts are applying formal operations to basic elements, or are following any rules or procedures at all. If a rule ("if X occurs, do Y until Z occurs") is seen as the basis for skilled performance, then how does an expert know when X and Z have occurred? Higher level rules must be invoked for this guidance, and still higher level rules are needed to guide the performance of hierarchy. (See Weitzenfeld, 1981a, for a fuller account of this problem.)

The work of Herbert Simon (Chase and Simon, 1973; Larkin et al., 1980) offers an alternative to a calculational theory. Simon and his co-workers have emphasized recognitional capacities. Their research raises the question of what role is played by recognitional vs. calculational capacities, for highly proficient personnel.

4. Research Rationale

The general goal of this research program has been to test the hypothesis that recognitional capacities account for expertise, and to extend our understanding of proficient performance.

We have developed and refined a theoretical description of proficient performance (Klein, 1980a) based on recognitional capacities. This description is currently at a level of specificity that allows empirical testing. Basically, our account contrasts proficient performers with novices, in terms of recognitional capacities, perceptual learning, and the availability and use of analogues.

4.1 Recognitional Capacity. We hypothesize a distinction between recognitional capacities and calculational capacities. Recognitional capacities allow a proficient performer to immediately recognize specific situations, and the relevance of goals and strategies. These place no strain on limited attentional or memory resources. Calculational capacities involve the use of working memory to examine contingencies. Using chess as an example, a grandmaster would display recognitional capacities in perceiving several pieces as one unit, or chunk. Chase and Simon (1973) have estimated that in the course of their experience, grandmasters acquire the ability to distinguish between approximately 50,000 patterns of pieces. Larkin et al. (1980) further propose that the ability to recognize and distinguish between larger sets of patterns is basic to proficiency development in a variety of domains. We suggest that the recognition of patterns is accompanied by a recognition of the types of reactions that are plausible in response to those patterns. Thus, a grandmaster should be able to recognize more plausible moves in a situation than would a lower-rated player. In contrast, calculational capacity consists of the ability to work out the implications

of actions. Continuing with our chess example, once patterns have been recognized, and plausible moves identified, analysis is needed in order to select the optimal move. This analysis would consist of the examination of moves, counter-moves, etc. Above a certain level of proficiency, we would expect these calculational capacities to remain constant. The only difference to expect is that grandmasters may have larger "chunks" to deal with, thus freeing more of their working memory space for deeper analyses. Glaser (1981) has discussed the importance of recognitional capacities over calculational capacities in accounting for proficient performance.

4.2 Perceptual Learning. Our second hypothesis is that qualitative differences in task perception influence success. Gibson (1969) has discussed the importance of perceptual learning for skills, and it seems obvious that experts can make distinctions that are opaque to novices. However, the perceptual learning obtained through experience may not simply result in smaller jnd's (just noticeable differences). We speculate that the expert has also learned <u>which</u> dimensions to use in examining tasks. That is, the learning necessary for proficiency is based on the acquisition of more powerful perceptual dimensions, as well as the ability to make finer discriminations along these dimensions.

These discriminations may reflect a greater ability to differentiate between appropriate goals. Sensitivity to overall goals leads to coordinated performance, as opposed to the jerky movements of novices reacting primarily to immediate demands.

4.3 Analogical Reasoning. A third hypothesis is that specific previous experiences can be used as analogues to a given problem situation, acting as an efficient means of bringing a large amount of information to bear

on the problem. Someone with more experience will have available a wider range of analogues, and will be likely to have available an analogue that is directly relevant to a specific task. In addition, a person with more experience will be able to make better use of analogues to define problems, generate options, anticipate outcomes, and formulate predictions.

However, it is difficult to test these hypotheses without a comprehensive theory of analogical reasoning. Accordingly, much of our effort in this domain has been in the direction of developing descriptive and prescriptive theories of analogical reasoning. The logic of analogical reasoning may govern the use of schemas and prototypes; it may serve as the basis for the process of generating new hypotheses. The study of analogical reasoning may have applied value if it can help us to understand how new situations are understood, and how predictions are arrived at.

4.4 Summary. We are assuming that expertise develops through perceptual learning rather than just through the acquisition of rules and procedures. Experts are not just faster and more accurate at applying rules and higher level rules. Rather, their skill is based on the fact that they have learned to perceive situations differently. They can perceive larger chuncks, and they can recognize overall situations and relationships. They can make discriminations that are opaque to personnel with less experience, and they can detect similarities that go unnoticed by personnel at lower skill levels. They have acquired a wide range of applicable experiences. Perhaps most important of all, experts appear to be able to recognize plausible goals in situations. They can examine a situation and quickly understand what sorts of outcomes are worth striving for. These goals appear to be recognized without the need for calculations. An expert simply seems to be able to recognize what out-

comes to expect, and what goals to emphasize. To some extent, this recognition may be based on perceived similarity of the current situation to other analogous situations, or to prototypes derived from several analogues. Once the expert has identified long-range goals, these can be used to structure short-range goals and plans. Thus, the performance of the expert appears smooth and coordinated because actions are generally occurring within a context of overall goals. In contrast, novices are usually reacting to local conditions, and trying to respond to immediate pressures. There are no long-range goals to integrate their performance.

4.5 Research Findings. We have studied recognitional capacities, perceptual learning, and analogical reasoning.

4.5.1 Recognitional Capacity.

Our prediction was that proficiency at a task depends on recognitional rather than calculational capacities. People who are more proficient at a task appear to be able to recognize better options and reactions, and this is a reason for their performance superiority. This prediction was tested in a study of chess expertise performed in collaboration with Stuart and Bert Dreyfus, University of California at Berkeley (who had been funded by the Air Force Office of Scientific Research Grant AFOSR-78-3594).

4.5.1.1 Subjects were three chess players rated as senior masters, and three players rated as experts. The U.S. Chess Federation rates players on the basis of outcomes of games and tournaments. A difference of 200 rating points translates into a 75% probability of winning a game. The median of American tournament players is estimated at 1400. Class E players (beginners) are rated less than 1200. Class D players are rated 1200 - 1399. Class C players are rated 1400 - 1599. Class B players are reasonably strong, and rated 1600 - 1799. Class A players are very strong, and are rated 1800 - 1999. The next step above Class A is the

rating of expert, 2000 - 2199. Masters are rated 2200 - 2399. Senior masters are rated above 2400. International grandmasters are rated above 2500, and in addition have shown a certain level of proficiency while playing in certified tournaments.

In this study, the three experts were rated 2062, 2130, and 2150. The three masters were rated 2401, 2403, and 2500. In addition, we used two players to rate the moves of the games played by the masters and experts. One of these players was rated 2500+ (this person was a senior master, lacking only tournament credentials to be considered an international grandmaster. At the time of the study, he had tied for first place in the U.S. Chess Championships held at Stanford in the summer of 1981). The other rater was rated at 2520 (international grandmaster).

4.5.1.2 Procedure. Two tournaments were arranged, one for the experts and one for the masters. Each tournament consisted of a double round-robin, in which each player played each other player two times, once with the black pieces and once with the white pieces. Regulation time of 50 moves in 2 hours was used. In addition, another double round-robin was played by the same players with only five minutes total available for each player. Thus, condition A was playing skill, master vs. expert, and condition B was time available, regulation time or a speeded condition.

For each set of players, the sequence of speeded and regulation games was counterbalanced for each session. Each player played white and black an equal number of times against each other player, and began a set of games with an opponent playing white and black an equal number of times. An equal number of sessions began with the 5-minute game first and the regulation time game first. This design yielded 6 regulation time games and 6 speeded games, at each of the two skill levels.

Incentives were provided for performance. For the experts, each player was paid $2.50 for each game played, and an additional $10 for each game won, whether regulation time or speeded. For the masters, there was $10 payment for playing each game, and an additional $30 for each game won; in addition, the results of each regulation game were presented to the U.S. Chess Federation, to be taken into account in revising ratings.

For each game played, records were made of the moves played. For the regulation games this was done during the games. For the speeded games the moves were reconstructed immediately after the game. A research assistant was present at all the games played by masters, and recorded moves during the speeded games. In addition, times were recorded along with moves during the games for the six regulation games played by masters.

After the games were played, sheets were prepared for the raters. Each game was coded, so there would be no indication of whether the game was regulation or speeded, played by experts or masters.

For each game played, moves 1-10 were deleted (since we weren't interested in studying knowledge of book openings), and a diagram was prepared with the position after move 10. Two chess players with ratings above 2500 were paid to rate each of the moves. The initial ratings were performed independently, but subsequent consultation was allowed to permit the sharing of discoveries about strengths and weaknesses of specific moves.

Each move was rated on two scales. First, the rater assessed the position prior to the move, and determined whether there was clearly one best move in that situation, or whether there were at least 2-3 moves to consider. The rationale for this rating was that the skill of the masters should be more evident in more complex situations. Second, the move selected was rated on a 5-point scale. The anchors for the scale are as follows: 5 (there are no moves better than this one), 4 (playable, but

not the best move), 3 (dubious, not a strong move, but not a blunder), 2 (a positional blunder, threatening the loss of material or an attack on the King), and 1 (a material blunder, leading to the outright loss of a piece). Ratings took approximately one hour/game.

4.5.1.3 Results. Agreement between raters was high. For the first scale (clearly one best move in a situation vs. at least 2-3 alternatives) there was 92% agreement. For the second scale, quality of moves,the correlation between the ratings for the two raters was approximately .84. The data were combined for the two raters by averaging the ratings given to each move. For the judgement of "clearly one best move" vs. "at least 2-3 alternatives," we only used the cases where both raters were in agreement.

The average game contained 40 moves by each player, of which we were able to use 99%. The overall data are presented in Figure 1. The four data points in Figure 1 each represent between 324 and 474 ratings. The rated quality of moves is higher for the masters than the experts, by .14. This difference was significant, $F(1,22) = 5.0$, $p < .05$. The difference may appear small, but it should be remembered that this is the average difference per move. Projected over a series of 7 moves, it would result in a master making one move rated as "5" while an expert was making a playable, but not highest quality move, rated "4." Projected over a game of 30-60 moves, the difference would be sufficient for the master to generally win games if matched with an expert.

The average time/move in regulation games was 2.5 minutes, and for speeded games it was approximately 6.0 seconds. Figure 1 also shows that the move quality was higher for regulation games than for speeded games, $F(1,22) = 17.1$, $p < .05$. This supports a calculational model, in which the players are constructing sequences of moves, counter-moves, counter-counter moves, and so on. The more time available to perform such

analyses, the deeper and more careful the analysis can be. Both the master and expert show improvement from speeded to regulation time games.

Figure 1 supports a mixed model, including recognitional and calculational abilities. If the superiority of the masters was due solely to calculational skills, then this superiority would be more strongly demonstrated for the regulation games than the speeded games. However, the trends for masters and experts are parallel. They both show the same improvement with time. These data do not support a model claiming that the superiority of the master is due to better tree construction and searching.

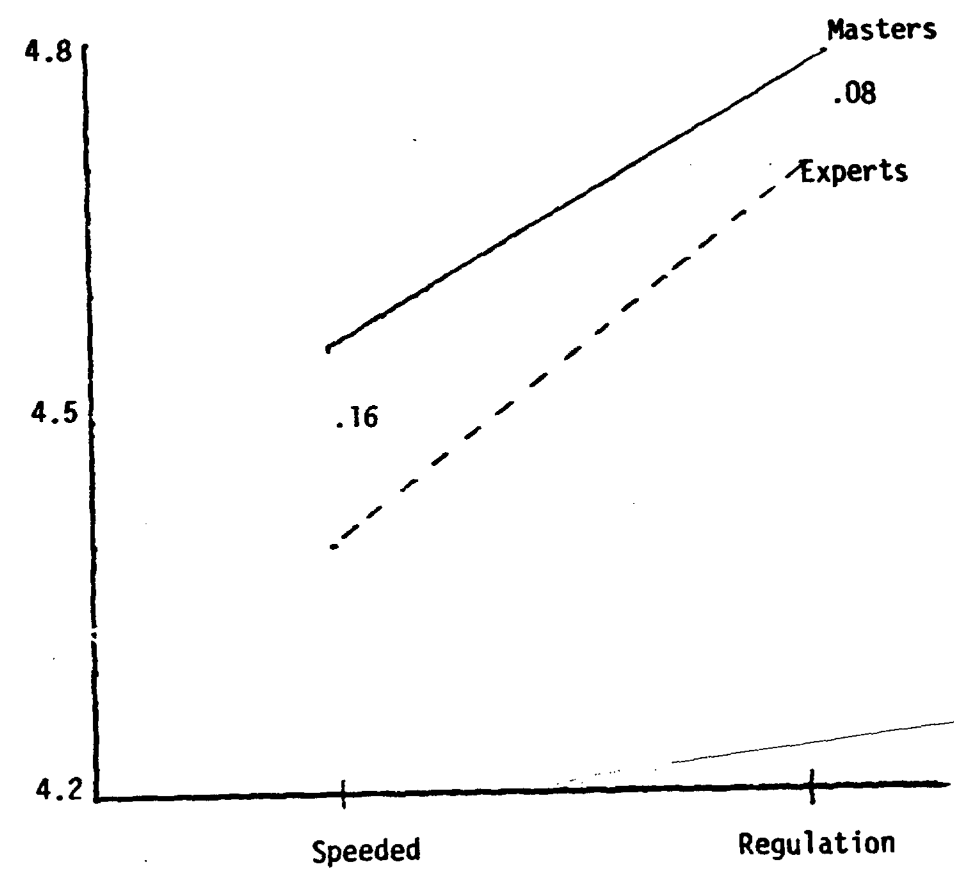
The data do support the type of recognitional model discussed by Simon. If the chess master can recognize better moves to analyze, we would expect this difference to emerge for the 5-minute games, as it does. The data are consistent with a model of chess decision-making in which a finite, limited number of moves are recognized, and are then analyzed. The master can recognize higher quality moves, but is not superior to the experts at analyzing the moves recognized. Thus, the difference that appears for 5-minute games remains constant for regulation games.

These data also support deGroot's (1978) observation that grandmasters could recognize and select the best move in a difficult chess problem and experts rarely even considered the move as an option.

The data were further analyzed into situations where there was clearly one best move (C), vs. situations where there were at least 2-3 good moves (2-3).*

* The validity of this distinction is supported by the time data for regulation games of masters. The mean time taken for situations where there was clearly one best move, was 1.68 minutes, whereas if there were at least 2-3 alternatives, the time taken was 3.99 minutes. The difference was significant, $F(1,23) = 35.37$, $p < .01$.

Figure 1: Move Quality for Masters vs. Experts, for Regulation and Speeded Games



These data are presented in Figures 2 and 3. Several findings emerge from examination of these figures. First, the parallel trends for masters and experts becomes even more pronounced. For Figure 2, the difference in move quality between masters and experts is .08 for regulation games, and .10 for speeded games. In Figure 3, the difference between masters and experts in move quality was .31 for regulation games and .26 for speeded games. None of these data can be interpreted as showing the masters are superior at calculations or analysis, compared to experts.

A second finding is that the difference between masters and experts is more pronounced when there is not any clearly best move. The slope of the lines in Figure 3 are also flatter than in Figure 2.

Since masters show a greater superiority to experts in complex situations (2-3 alternatives) than in simple situations (one best move), we would expect that this would affect the strategies used by each type of player. Figure 4 shows that under speeded conditions, both masters and experts show the same proportion of cases for which there are 2-3 options: 42% of the total moves. Given enough time to shape strategy, masters reduce their proportion to 36%. However, the experts reduce the proportion to only 18%. The experts seem to be trying to maximize the role of their calculational skills, and minimize their reliance on recognitional skills. They are simplifying their games. It would be interesting to compare these proportions for games in which masters and experts faced each other.

Figure 4 also explains why the trends in Figure 1 are not as parallel as those in Figures 2 and 3. The experts were playing more simplified regulation time games than the masters (in more than 80% of the cases there was one clearly best move), and the average rating for such moves was higher

Figure 2: Move Quality for Situations Where There is Clearly One Best Move

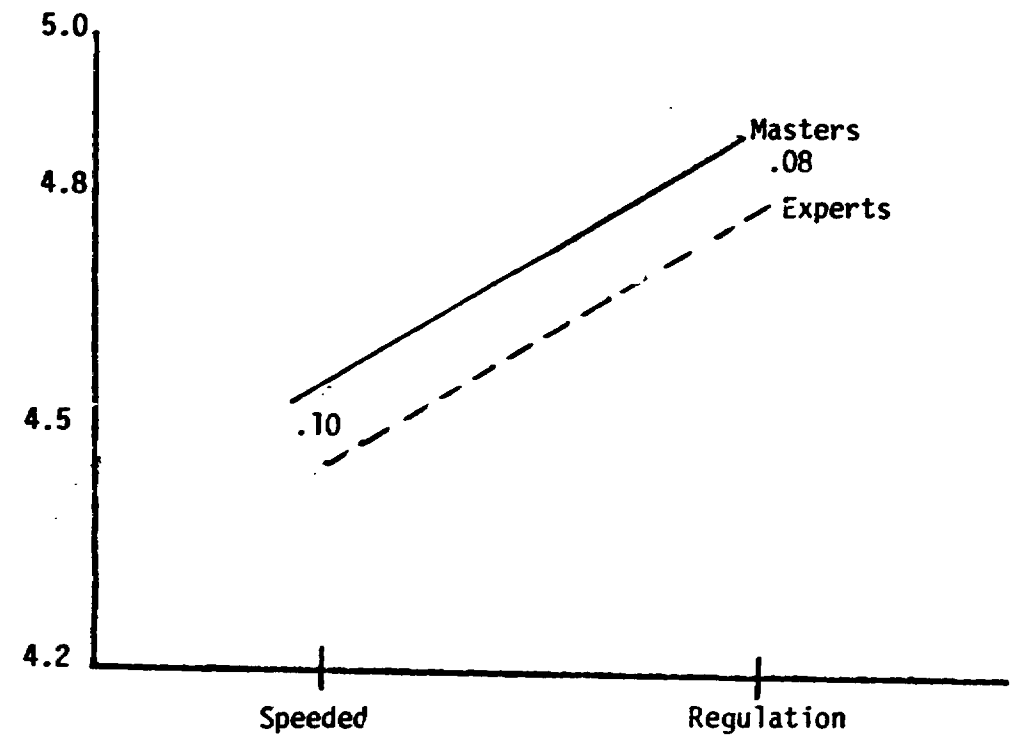


Figure 3: Move Quality for Situations Where There are At Least 2-3 Alternatives

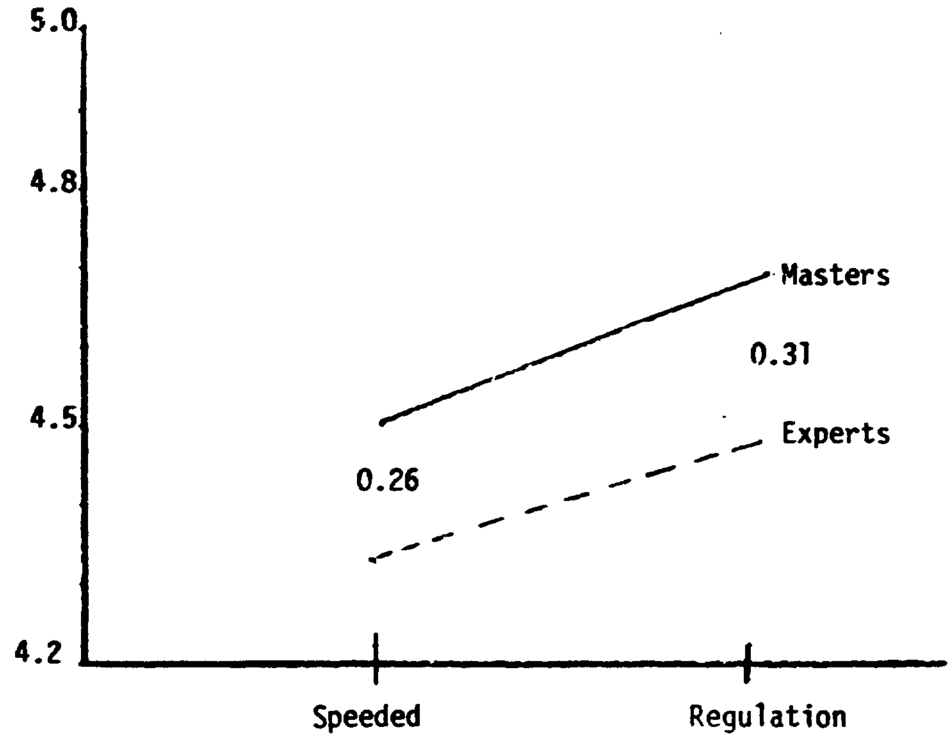
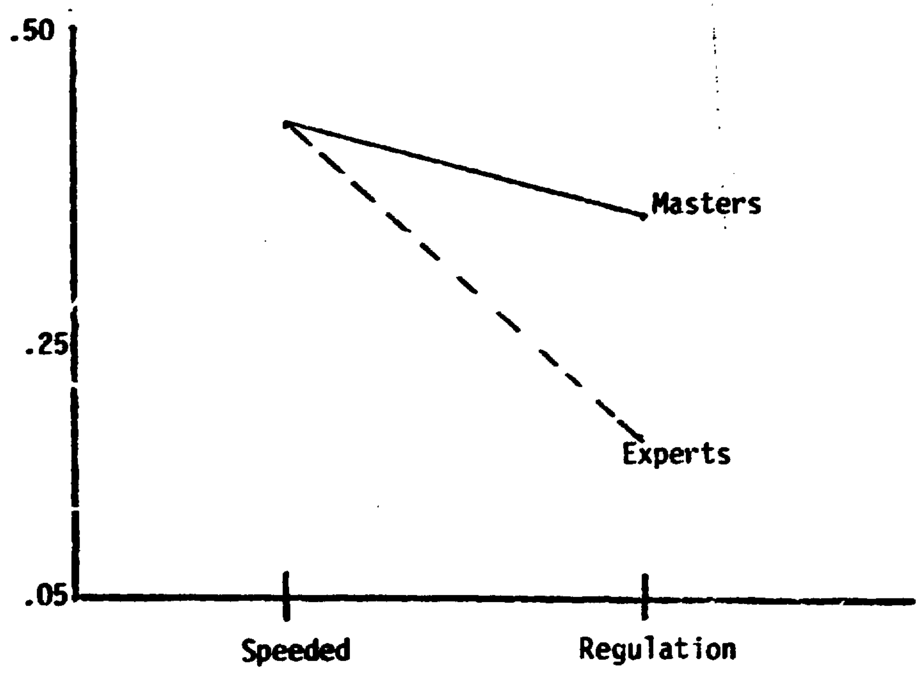


Figure 4: Ratio of moves with 2-3 alternatives to total moves



than for situations with 2-3 alternatives. Therefore, the data for expert moves in regulation games are artificially inflated in Figure 1. Figures 2 and 3 maintain the distinction between the C condition and the 2-3 condition, and are a more accurate reflection of performance.

The data should not be interpreted to mean that skill level is never related to calculational skills. We only examined two neighboring classes of skills, both at very high levels. We would expect that at lower skill levels, calculational capacities would emerge as a differentiating factor.*

4.5.1.4 Implications. The data support the importance of recognitional capacities for highly proficient performance. This suggests that it may be more fruitful to study high levels of proficiency in terms of perceptual learning models than in terms of tree-searching, calculational models.

The data have implications for training. The training of recognitional capacities needs to be examined if we are to be able to use training programs to develop high levels of expertise. Simon and Chase, and Larkin et al. claim that such capacities are developed only after thousands of hours of practice. Highly proficient subjects appear to be able to distinguish between 50,000 different patterns. The pattern vocabulary for good club chess players is only about 1,000 patterns. Novices can only recognize a few patterns. One challenge is to be able to expand the recognitional pattern vocabulary more efficiently. In chess this might be developed in beginners by developing

* However, it is difficult to obtain ratings for moves at low levels of play, for several reasons. The ratings would be biased, since it would be clear to raters that skill levels were markedly different. In addition, it would be difficult to rate moves, since a given move might be a blunder committed by a Class C player, or a clever tactic played by an expert. The rater would have to see how the move was followed up in order to determine how much strength to read into it, and this would complicate the ratings.

training materials consisting of games played by high level players. The task would be to predict which moves in each situation were considered by the players; these predictions would be matched against an actual listing of the moves that the players did consider. Such training could facilitate the ability to recognize the types of moves to be examined.

The data also have implications for the design of decision aids. Two inferences are made. The first is that in order to help experts play like masters, they will need to have a better set of initial moves to consider. Second, since both masters and experts are relying on calculational skills, a decision aid that allowed the player to enter initial moves (and counter-moves) and then performed the subsequent tree construction and search, might be of benefit. The interaction would consist of the operator continually pruning the decision tree by rejecting poor lines of play, and emphasizing promising lines for deeper analysis.

Finally, the results have implications for workload assessment. When the average time/move is reduced from 2.5 minutes to 6 seconds, the quality of move is reduced by only a small amount, and is still reasonably high. For the speeded games, the average move generated by experts was still rated above the level of playable. This suggests that for tasks that involve recognitional capacities, and are performed by proficient personnel, time pressure may not have an overwhelming effect on performance. It must be remembered that this holds, in the present experiment, for players rated as experts, who are far inferior to grandmasters.

In fact, it could be argued that the skill of the players studied is primarily recognitional, rather than calculational. This argument, which depends on several tenuous assumptions, runs as follows: If we assume that recognitional capacities are manifested within the first few seconds, and

not thereafter (which is unlikely; evaluation of moves depends on recognition), then all of the improvement between the speeded and the regulation games is due to calculations. This is not very much improvement. Further, if we assume that only a minimal amount of calculations can occur within 6 seconds, then we might conclude that most of the skill depends on recognition, because very high quality moves were being generated within a very short time. Proponents of a calculational model would have to show that moderately good chess players are able to perform the analyses necessary to generate playable moves in only a few seconds. The burden of proof is on such proponents.

Of course, it must be noted that the paradigm we used did not control time available. For speeded games, the average time available was approximately 6-7 seconds. Subjects were undoubtedly using more time for more complex situations. They were also using analyses developed during prior moves, plus analyses performed during the opponent's turn. A design that provided better controls on time would be a next step for this research.

4.5.2 Perceptual Learning. The primary method we have developed for contrasting the perceptual abilities of experts vs. novices is based on similarity and difference judgements (Galanter, 1956; Fransella & Bannister, 1977). The paradigm has three stages: (a) the selection of representative examples; (b) the elicitation of similarity/difference judgements, using those materials, to identify dimenisons of analysis; (c) presentation of rating scales to subjects at different levels of competence, to identify commonalities in the use of some dimensions, and to highlight dimensions that are used differentially. This paradigm allows us to determine perceptual differences between an expert and a novice. It allows us to measure processes

like the perceptual learning discussed by Eleanor Gibson (1969) for finding the relevant information in a situation. We think that experts use different discriminative dimensions than novices. This prediction was tested in a study that compared Cardio-Pulmonary Resuscitation (CPR) performance of students, CPR instructors, and paramedics.

4.5.2.1 Method

4.5.2.1.1 Subjects. Three groups of subjects were used in this study: students, instructors and paramedics. The students were adults who had taken a CPR training program (eight hours) and had received certification for successful completion. The instructors had completed the CPR training program as well as an additional training program for instructors. Each instructor had received instructor certification at completion, and had taught CPR to novices. No instructor in this study had ever actually performed CPR on a victim. The paramedics were trained in CPR and had experience using CPR with victims as part of their work. None of the paramedics in this study had been involved in presenting CPR instruction.

4.5.2.1.2 Materials. The study used videotapes and test booklets. There were six different videotapes each showing a person (exemplar) doing cardio-pulmonary checks and performing CPR on a ResusciAnne training simulator.

The test booklet presented the judgement dimensions that had been derived during an earlier study of CPR expertise and it provided a place for the subjects to enter a rating for each videotaped performance for each of 13 dimensions. The subjects also indicated which of the six people they would choose to save their own life in an emergency.

4.5.2.1.3 Procedure. Each subject saw videotape presentations of

the six exemplars. The task was to rate each of the six performances on each of the 13 dimensions. For each dimension, both ends of a scale were described, e.g., "Smooth" and "Jerky." One end of the scale was to be described as "1" and the other as "5." Each videotaped performance was to be rated between "1" and "5" for each of the dimensions.

4.5.2.2 Results

How did each of the three groups judge exemplar skill?

All subjects were asked which exemplar they would choose to save their own life in an emergency. (The six tapes showed five students and one paramedic.) In the paramedic group, 9/10 subjects selected the CPR performance of the paramedic exemplar. The paramedic exemplar was selected by 5/10 students, and by only 3/10 instructors. The instructors were concerned that the paramedic wasn't following the procedures they taught in their courses.

The judgement pattern on individual dimensions was consistent with this finding. The paramedic group judged the performance of the paramedic exemplar highest on 12 of the 13 dimensions ("hand placement" was the only exception). The student group judged the paramedic exemplar highest on only 6 dimensions: "smoothness," "compressions simulate heart action," "efficient," "compression time correct," "confident," and "performance reflects an understanding of how the body works." The instructor group judged the paramedic exemplar highest on only 4 dimensions: "smooth," "adequate breath check," "correct pulse assessment," and "performance reflects an understanding of how the body works."

How do the groups differ in their use of dimensions?

In general, the paramedics were able to use all the 13 dimensions

to discriminate between the performance of the exemplars, whereas the instructors and novices had difficulties in using several of the dimensions.

Differences between groups taken two at a time were examined for each of the 13 dimensions. Table 1 presents *p* levels of significant single discriminant functions. For student vs. paramedic judgements, the single discriminant functions were significant for six dimensions. Students and instructors could be significantly distinguished with the single discriminant functions for ten of the dimensions. Finally, the judgement of the instructor and paramedic groups could be distinguished on nine dimensions. As can be seen in Table 1, there were significantly different patterns of judgements between all three groups for five of the 13 dimensions.

Differences were found between groups in the use of specific dimensions. An example of a large difference in the use of a dimension is shown in Figure 5. Figure 5 presents the performance patterns for instructors and paramedics, for the dimension of "efficiency." The mean rating for each exemplar is given, along with the band size of one standard deviation. Exemplar A is the videotape for the paramedic; the other five exemplars are videotapes of students performing CPR. Figure 5 shows how the paramedics could use this dimension to distinguish the paramedic from the students, whereas the instructors were primarily distinguishing between students.

4.5.2.3 Discussion

The results support the hypothesis that personnel at different skill levels will show differential use of the same dimensions in perceiving performance of a task. The paradigm that was used is capable of showing which dimensions were consistently used by which group of subjects, as well as identifying cases in which two groups of subjects were both using a dimension, but in different ways.

TABLE 1

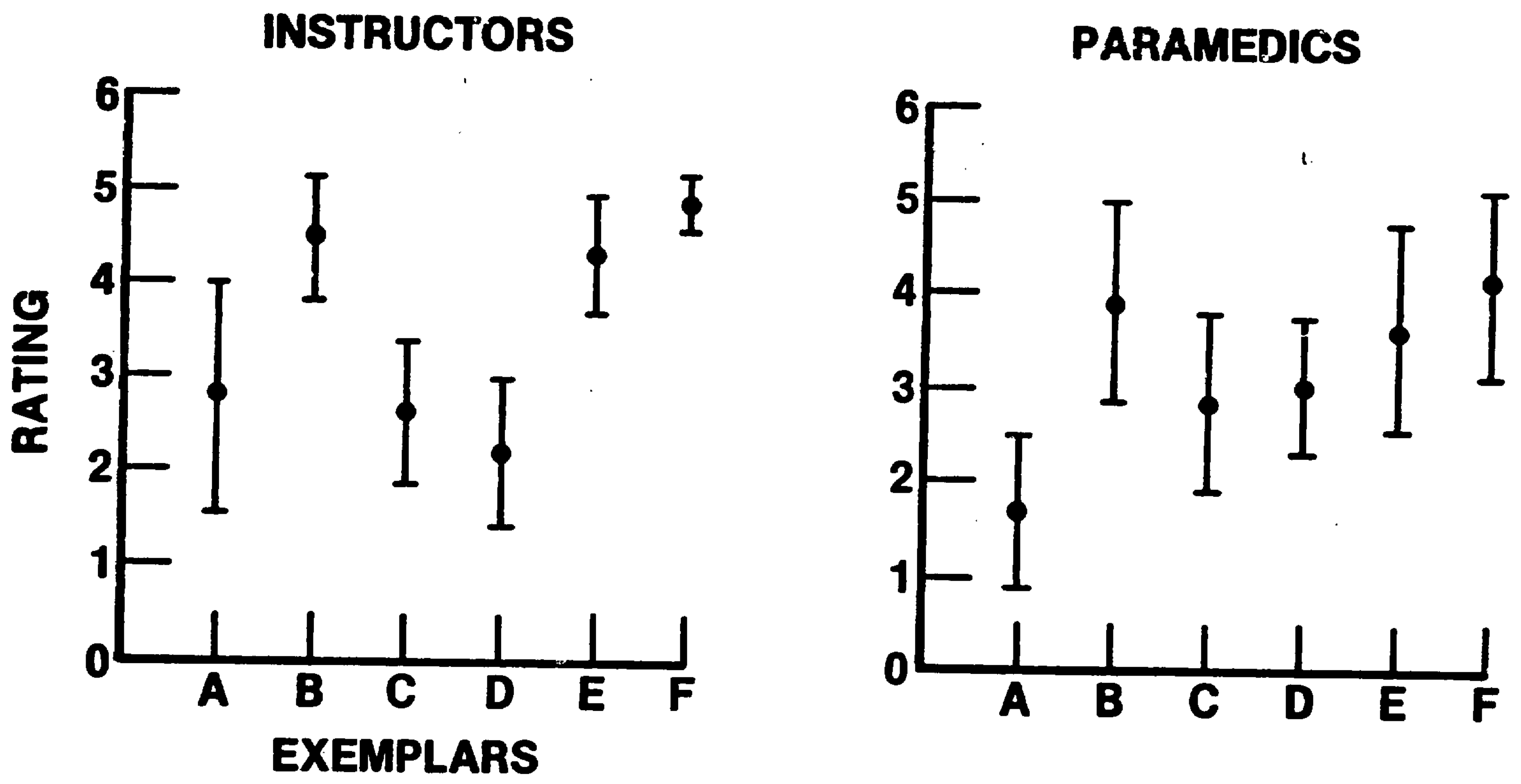
Significant levels for single discriminant functions,
for the three pairs of groups

DIMENSION	Student/ Paramedic	Instructor/ Student	Instructor/ Paramedic
1. Smooth-Jerky	*	*	*
2. Compressions simulate heart action - compressions fail to simulate heart action		*	
3. Compressions timing correct - timing incorrect			**
4. Dangerous - Effective	*	*	*
5. Checks cues (monitor) - fails to check cues		*	
6. Body position over victim - body position at right angle	*	**	
7. Hand placement acceptable - hand placement unacceptable			**
8. Adequate breath check - inadequate breath check		**	*
9. Correct pulse assessment - incorrect pulse assessment	**	**	*
10. Efficient - inefficient	*	**	**
11. Confused - confident		**	*
12. Compression depth correct - compression depth incorrect	*	*	*
13. Performance reflects understanding of how body works - performance reflects ignorance of how body works			

.05*

.01**

Figure 5: Instructors vs. paramedics in the rating of efficiency



The research paradigm can be used to make decisions about training requirements. This is especially important at higher skill levels, where proficiency depends on the way people have learned to perceive situations, rather than on rules and procedures. The results of Table 1 can be used to identify training requirements for instructors, to help them learn to perceive the CPR task more like the paramedics do For nine of the thirteen dimensions, there was a significant difference in the way instructors and paramedics perceived the exemplar performance. Three of these differences were significant at the .01 level.

The general strategy for identifying training requirements is as follows: for any given dimension, are the more skilled subjects using that dimension? If not, then it does not require training. If so, then we must see if the less skilled subjects are also using the dimension. If not, then it is a training requirement. If they are, but not in the same way as the more highly skilled subjects, then it is also a training requirement and the differences can be called to their attention as a training method.

In addition to identifying training requirements for more proficient personnel, the analysis of perceptual dimensions may also have some value in evaluating training progress, and in supporting personnel selection decisions. Applicants can be matched to existing group profiles, to see which group is most closely matched by their perceptual judgements.

The research paradigm appears to have general applicability to domains in which there is a contrast between novice and proficient performance. It is currently being used to study computer programming skills (Peio, 1981).

Computer software experts (more than seven years experience) were contrasted with novices who had completed high levels of training or had one to three years experience in the field. As in the CPR study, relevant task dimensions were elicited and subjects ranked task exemplars on the basis of those dimensions.

The task selected was the evaluation of algorithms for a particular programming task. A variety of algorithms for a classical computer programming task, that of the shortest path problem (critical path analysis) were chosen for the study. These algorithms differed in a variety of ways, and using the matching technique ten dimensions were elicited. Subjects were then asked to examine these exemplars and rank them on the basis of how the dimensions apply to each one.

A discriminant analysis again revealed significant patterns of differences between groups of experts and novices in the use of perceptual dimensions. These dimensions successfully separated novice and expert groups on seven of the ten dimensions, $p < .05$. Furthermore, the discriminant analysis revealed which individuals were correctly classified as experts or novices solely on the basis of scores on those dimensions. These predictions ranged from 75% to 95% correct on the seven significant dimensions. In four of the most significant dimensions (over 85% correct classifications) the same novice programmer accounted for 5% of incorrect classifications of group members. It was later found that this novice was experienced with this particular programming task. An additional finding was that experts were able to use more of the dimensions than the novices in discriminating between algorithms, providing further support for the usefulness of this paradigm in determining training requirements and in the identification of proficient personnel. (See Table 2.)

TABLE 2

Group Significance of Dimensions

Dimensions	Significance	Percent Correctly Classified+
Independence of Computer Strengths	** 0.008	95.0%
Readability	** 0.001	90.0%
Writability	** 0.005	89.5%
Subpaths	** 0.002	75.0%
Storage	* 0.015	85.0%
Language	* 0.036	80.0%
Execution Time	0.070	70.0%
Validation	* 0.047	80.0%
Nodes	0.077	65.0%
Calculates Paths	0.165	70.0%

* $p < .05$

** $p < .01$

+ Percent correctly classified reflects group members who were correctly predicted as being expert or novices by discriminant analysis SPSS version eight, Nie, Hull, Jenkins, Steinbrenner, and Bent, 1975.

4.5.3 Analogical Reasoning. Our goals in this domain have been to develop descriptive and prescriptive models of analogical reasoning within the context of generating predictions in technological environments.

A large body of previous work (e.g., Sternberg, 1977) on analogy has focused on the four-term judgement format, a:b::c:d. This format is very common in educational measurement. We feel that no existing model of this format can be applied to any other reasoning by analogy, since the four-part analogy problem does not require subjects to identify or select analogues, or use analogues to solve problems. The four-part analogy problem primarily tests a subject's ability to recognize analytical and cultural factors that make certain types of similarity more relevant than others.

We felt that the analysis of analogical reasoning presented by philosophers of science (e.g., Hesse, 1966; Kuhn, 1962) would be more applicable to technological environments. These arguments were presented by Weitzenfeld and Klein (1979).

We tested the two approaches (four-part analogy problem vs. philosophy of science model) in a study with Air Force engineers. We interviewed seven engineers who had participated in an effort to predict the reliability of subsystems for the B-1 aircraft. The method they use, comparability analysis, consisted of comparing analogous subsystems on aircraft currently in operational use. Essentially, they were reasoning by analogy in a technological domain. Our interviews attempted to learn how they were doing this. Our results (Klein and Weitzenfeld, 1980) did not support either of the two approaches we were testing. We found that the Sternberg model simply was not relevant for the main activities of the engineers: selecting, rejecting, modifying, and using comparison cases. However, the philosophy of science model was also inadequate, because the

concerns of science are different from those of technology. The scientist wishes to identify new hypotheses to test; accordingly the philosophy of science model emphasized the ambiguous features of the two analogues, (features which were not clearly similarities or differences) and regarded them as sources of new hypotheses. The goal in technology is to predict specific items of information, not to discover new hypotheses. Therefore, the engineers were attempting to identify comparison cases that allowed them to predict the reliability of specific B-1 subsystems.

By examining their strategies, we obtained a clearer understanding of the task of technological prediction. The use of analogies or comparison cases has always seemed risky in such situations, because the general feeling has been that the force of the comparison is based on the extent of similarity between the target and comparison domain, and is therefore probabilistic. Our attention was turned to the rational basis for reasoning by analogy. We found that there was no sound basis for drawing inferences on the grounds of degree of similarity (Weitzenfeld, 1980). We do not think people actually reason that way. Instead, we believe that analogical reasoning is based on deductive, rather than probabilistic, reasoning.

Weitzenfeld (1981) has been able to define the necessary conditions for obtaining valid inferences using analogical reasoning. This paper is important for several reasons. First, its new deductive rationale for analogical reasoning is radically different from previous accounts. (It bears some resemblance to work in the cognitive sciences, but none to philosophical or psychological models.) It explains both why reasoning by analogy is so common (it can be as valid as deductive reasoning) and why it fails so often (it requires premisses that are hard to establish).

Second, by specifying the conditions under which such reasoning is

valid it raises cautions about inferences from comparisons that do not satisfy these presuppositions. Third, it is the basis for a prescriptive account of how to go about using analogies. We are now applying it to problems of training device design. Fourth, it can be the basis for descriptive theories of human use of analogues. The data we elicited from engineers fall into place when interpreted as intuitive applications of this model.

We think the study will be important in the philosophy of science and the psychology of analogy, as well as being a normative tool. It shows that the use of comparisons by experts must be a more complex process than has been thought if it is to lead to valid conclusions.

The new account of reasoning by analogy is based upon identities of structure among systems. It discusses the variety of forms of structure and their relative stability. It provides a formal definition of structural identity that avoids difficulties encountered by previous definitions. It discusses ways of discovering the existence of such identities and shows how they license different inferences. Among the conclusions is the central methodological rule for analogy: select analogues to match on variables that are not understood and then correct for the differences that you do understand.

All of this work, including the initial model, the research with the engineers, the analysis of similarity, and the prescriptive model, was supported under the present contract. This work is currently being continued in efforts funded by the U.S. Army Research Institute, to predict the training effectiveness of new training devices.

4.5.4 Decision Making. Our work in decision making was essentially an outgrowth of our research into proficient performance. The domain of interest was tactical C^2 decision making. Our hypothesis is that expertise at such decision making depends on recognitional capacities and analogical

reasoning, developed through many hours of experience. We are skeptical of calculational approaches to such decision making. Therefore, we were interested in attempts to develop automated decision aids in this area. We feel that such aids can represent signficant improvements in efficiency and performance. However, we are concerned that such aids could diminish the performance of proficient decision makers if the aids are based upon a calculational model of proficiency and so prevent the skilled decision maker from utilizing recognitional capacities. These issues were described in a working paper (Klein, 1978) and were presented at two conferences (Klein, 1980b; 1981).

5. Research Applications. In addition to providing a test of a general theory of proficient performance, the experiments described above may have implications for a variety of applied issues.

5.1 Decision Making. The general area of the development of automated decision aids for tactical C^2 presupposes that complex tasks can be divided into basic elements, and that decision analytic procedures can be applied as formal operations, to provide guidance for command battle managers. However, the perceptual/recognitional description of proficient performance that we have been developing raises some questions about such work (Klein, 1980b). We have attempted to use this work to derive guidelines for the allocation of decisions within the human-system interface, in a way that is consistent with the skills of the experienced operator.

5.2 Predictive Logic. Requirements to predict reliability of subcomponents of new aircraft, or to predict training effectiveness of new simulation devices, seem to be based on reasoning by analogy. As we gain a clearer understanding of how people identify and use analogues, we should be able to provide more specific guidance for tasks requiring predictions. The analysis developed by Weitzenfeld (1981b) presents a prescriptive model for the activities required in order to ensure valid predictive capabilities.

5.3 Training Requirements. Simple rule-based descriptions of tasks are usually sufficient for defining training requirements when dealing with novices, and with procedural tasks. However, when dealing with non-procedural tasks, and with developing higher levels of proficiency, new methods are needed for identifying training requirements. The research reported should be valuable in this effort, by demonstrating the relevance of perceptual

learning, goal frameworks, and analogical reasoning for capturing the basis for proficient performance. This offers the possibility of deriving training requirement analyses using methods for gauging the perceptual dimensions used by trainees, the sophistication of their recognitional capacities, and the types of analogues that they have available for use.

5.4 Workload. Our research with proficient chess players demonstrated their ability to maintain competent performance under extreme time pressures. Presumably, recognitional capacities are not as affected by limitations on working memory as calculational capacities. This raises questions about the ability to use recognitional strategies to overcome workload requirements.

6. References

Chase, W., & Simon, H. A. The mind's eye in chess. In W.G. Chase (Ed.) Visual information processing. NY: Academic Press, 1973.

deGroot, A. D. Thought and choice in chess (2nd Edition). NY: Mouton, 1978.

Fransella, F., & Bannister, D. A manual for repertory grid technique. New York: Academic Press, 1977.

Galanter, E. H. An axiomatic and experimental study of sensory order and measure. Psychological Review, 1956, 63, 16-28.

Gibson, E. J. Principles of perceptual learning and development. New York: Meredith Corporation, 1959.

Glaser, R. The future of testing. American Psychologist, 1981, 36, 923-936.

Hesse,, M. Models and analogies in science. Notre Dame: University of Notre Dame Press, 1966.

Klein, G. A. The analysis of highly proficient performance. Paper presented at the American Psychological Association meetings, Toronto, 1978.

Klein, G. A. Task perception and proficiency. Working paper, 1980a.

Klein, G. A. Automated aids for the proficient decision maker. Paper presented at the 1980 International Conference on Cybernetics and Society, Boston, October, 1980b.

Klein, G. A. A perceptual/recognitional model of decision making. Paper presented at Summer Computer Simulation Conference, Washington, DC, July, 1981.

Klein, G. A., & Weitzenfeld, J. The use of analogues in comparability analysis. Unpublished manuscript, 1980.

Kuhn, T. S. The structure of scientific revolutions. Chicago: University of Chicago Press, 1962.

Larkin, J., McDermott, J., Simon, D. P., Simon, H. A. Expert and novice performance in solving physics problems. Science, 1980.

Nie, N. H., Hull, C. H., Jenkins, J., Steinbrenner, K., & Bent, D. H. SPSS statistical package for the social sciences, second edition. New York: McGraw-Hill, 1975.

Peio, K. Repretory grid analysis of skilled performance: software specialists. M.A. Thesis, 1981, in progress.

Sternberg, R. J. Component processes in analogical reasoning. Psychological Review, 1977, 84, 353-378.

Weitzenfeld, J. The scientific study of cognition. D. Phil. Thesis, Oxford University, 1977.

Weitzenfeld, J. Similarity and purpose. Unpublished manuscript, 1980.

Weitzenfeld, J. Knowledge-how and natural kinds in psychology. Paper presented at the New Jersey Regional Philosophical Association, Princeton, 1981a.

Weitzenfeld, J. Valid reasoning by analogy. Submitted to The Journal of Philosophy, 1981b.

Weitzenfeld, J. & Klein, G. A. Analogical reasoning as a discovery logic. Klein Associates Technical Report TR-SCR-79-5, 1979.

7. Cumulative Chronological List of Written Publications in Technological Journals.

Klein, G. A. User guides: Some theoretical guidelines for their use. Submitted to Instructional Science.

Klein, G. A. and Weitzenfeld, J. The use of analogues in comparability studies. Submitted to Applied Ergonomics.

Weitzenfeld, J. Valid reasoning by analogy. Submitted to The Journal of Philosophy.

Klein, G. A. and Dreyfus, S. Recognitional vs. calculational capacities in highly proficient performance. In preparation.

8. Professional Personnel Associated with the Research Effort.

Gary A. Klein, Ph.D. Principal Investigator

Julian Weitzenfeld, Ph.D.

Karen Peio (Master's Thesis currently underway, anticipated Fall, 1981)

9. Interactions

(a) Klein, G. A. User guides: guidelines for their use. Paper presented at APA convention, New York City, 1979.

(b) Presentation on analogical reasoning; Air Force Human Resources Laboratory, Wright-Patterson AFB, Ohio; July, 1979.

(c) Klein, G.A. Automated aids for the proficient decision maker. Paper presented at IEEE Conference, Boston, MA, 1980.

(d) Klein, G. A. A perceptual/recognitional model of decision making. Paper presented at Summer Computer Simulation Conference, Washington, DC, July, 1981.

(e) Presentation on highly proficient performance; Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio; July, 1981

(f) Klein, G. A., & Klein, H. A. Perceptual/cognitive analysis of proficient cardio-pulmonary resuscitation (CPR) performance. Paper presented at the Midwestern Psychological Association meetings, Detroit, Michigan, 1981.

(g) Weitzenfeld, J. Knowledge - how and natural kinds in psychology. Paper presented at the New Jersey Regional Philosophical Association, Princeton, 1981.